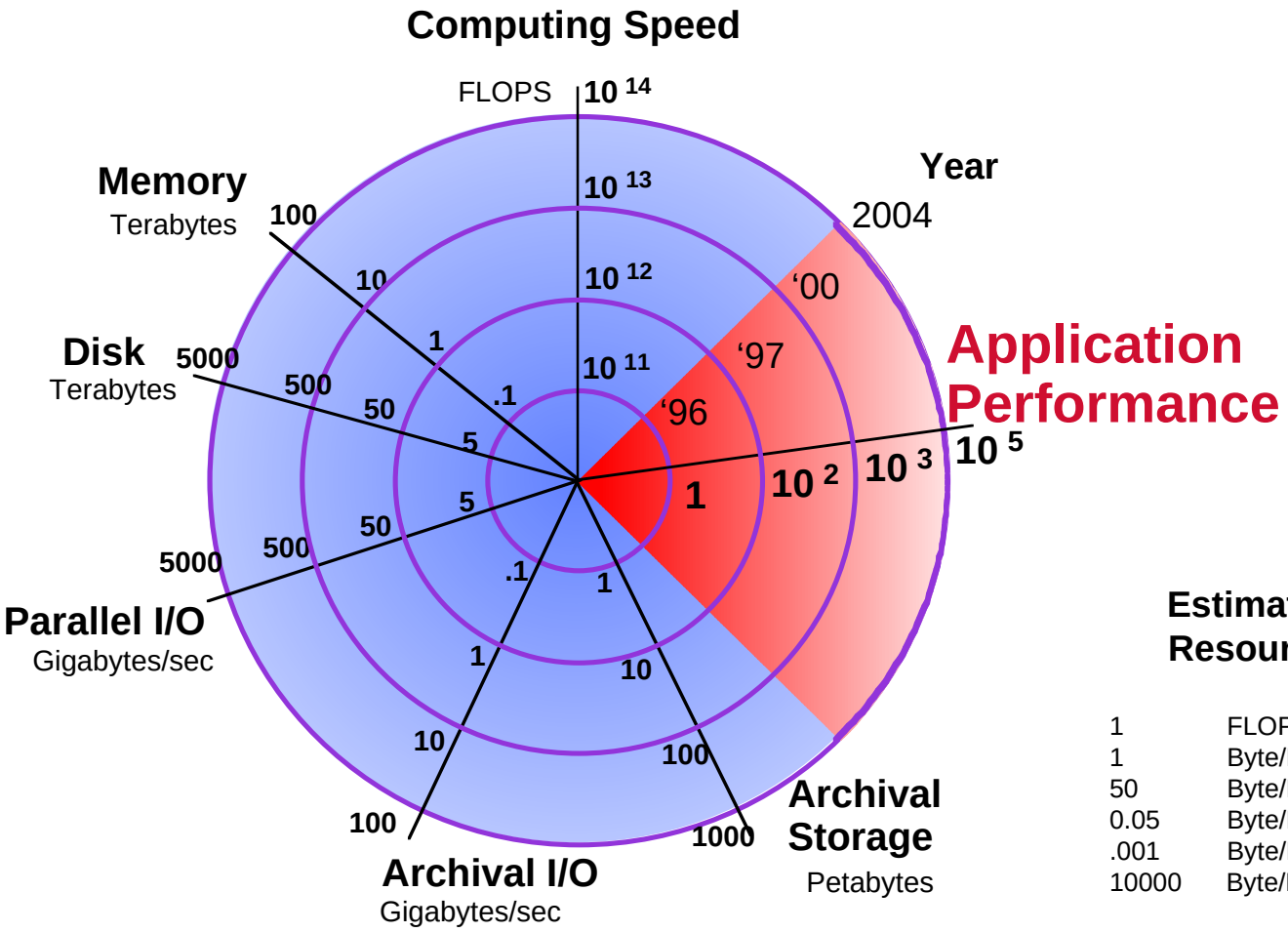
The Key to a Usable System is Application Driven Scaling

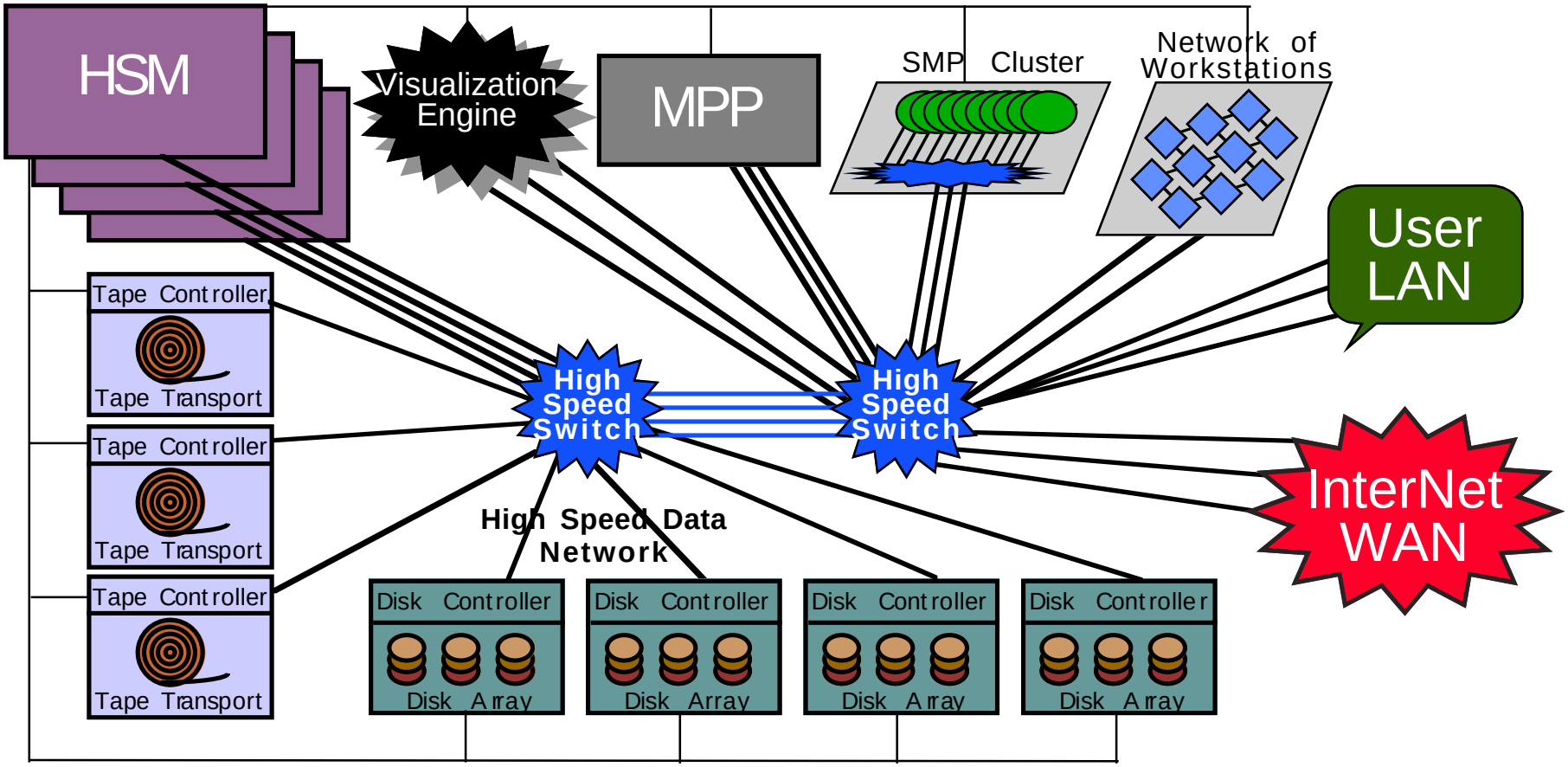


Estimated Computational Resource Scaling Factors

1	FLOP/s	Peak Compute
1	Byte/FLOP/s	Memory
50	Byte/FLOP/s	Disk
0.05	Byte/FLOPS/s	Peak Disk Parallel I/O
.001	Byte/FLOP/s	Peak Parallel Archive I/O
10000	Byte/FLOP/s	Archive

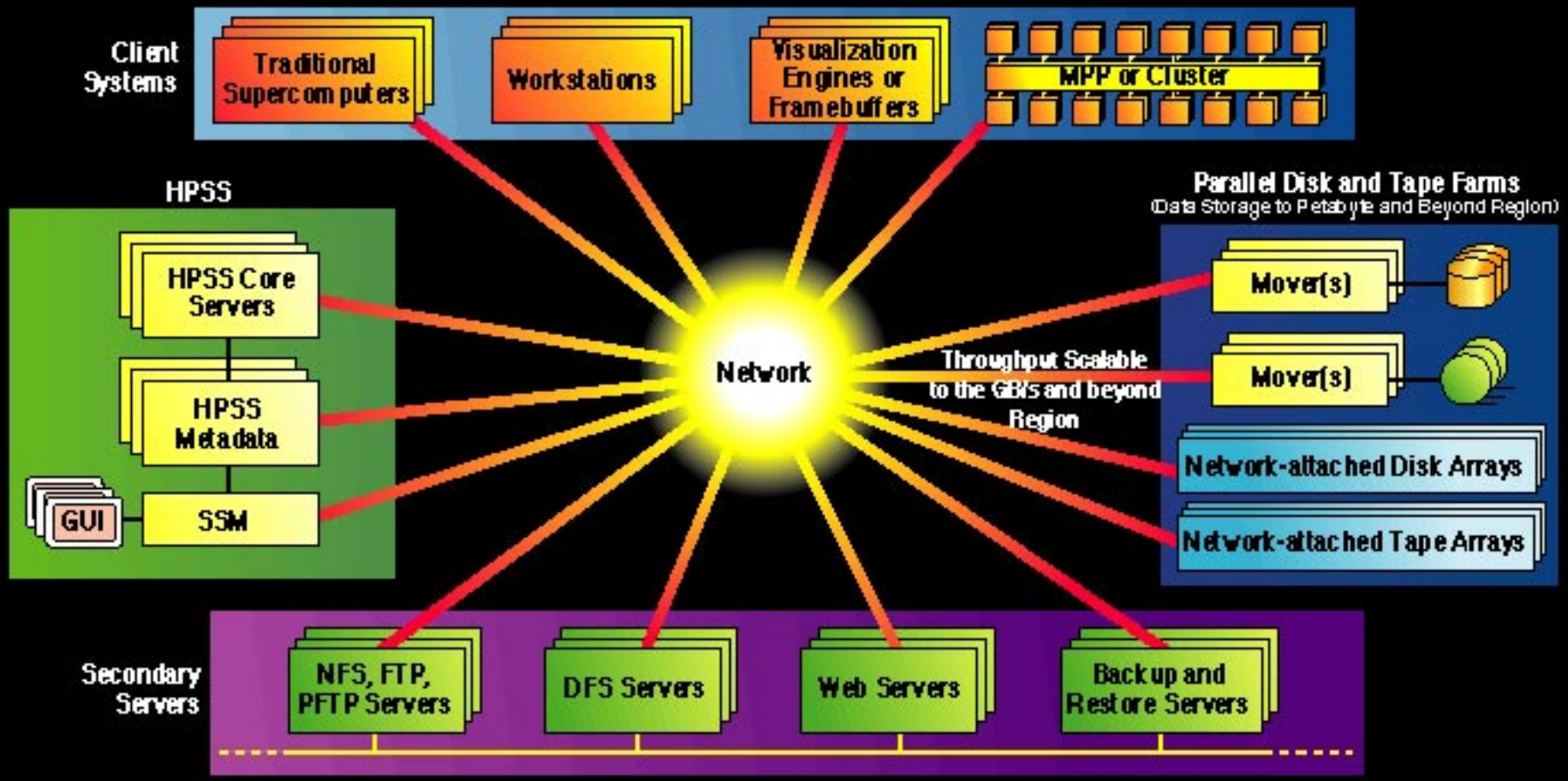
Systems = Scalable Network

THE NETWORK IS THE COMPUTER!!!



System Architecture Supported by HPSS

HPSS



Issues

- **Cost- we need new architectures to archive scalibility goals integrating commodity disk and tape (or equivalent) components and systems.**

- **I/O Bandwidth**
  - **Parallel tape systems (RAITSwith parity) and associated parallel robotics.**
  - **Wider RAIDS or equivalent.**
  - **Network (or interconnection fabric) attached peripherals (NAPs).**

- **Footprint**
  - **Media density or packaging improvements**

ASCI hardware requirements

Level	Effective Latency (CPU cycles)	Bandwidth (Random read/write)	Size	
On-chip cache**, L1	2-3 ●	16-32 B/cycle ●	10^{-4} B/flop * ● ↑	
Off-chip cache**, L2 (SRAM)	5-6 ●	16 B/cycle ●	10^{-2} B/flop * ● ↑	
Local main memory (DRAM)	30-80 (15-30) ↓	2-8 B/flop pk (2-8 B/flop sustained) ↓	1 B/flop ● ↑	Compute engine
"nearby nodes"	300-500 (30-50) ↓	1-8 B/flop (8 B/flop) ↓	1 B/flop ●	Interconnect
"far away nodes"	1000 (100-200) ↓	1 B/flop (1 B/flop) ↓	1 B/flop ●	
I/O (memory disk)	10 ms ●	0.01-0.1 B/flop ●	10-100 B/flop ●	
Archive (disk-tape)	Seconds ●	10^{-5}-10^{-4} B/flop (0.001-0.01 B/flop) ↓	10^{2} B/flop 10^{4} B/flop ↓	
User access	1/10 s (1/60 s)	OC3/desktop (OC12-48 /desktop) ↓	100 users ●	
Multiple sites	1/10 s ●	●	●	

* Equivalent integer and floating-point data calculation rates are required.

** Cacheless systems with equivalent performance are fully acceptable.

Primary investment priority

Secondary investment priority

1996-1998 Situation

(1998-2000 Requirements)

Industry Trend

↑ Industry gets better at meeting requirements

↓ Industry gets worse at meeting requirements

● Industry continues to meet requirements

Technical Goals

(Order of Magnitude Required for ASCI)

	2001	2004	2007
• Computing Power	• 20 TFLOPS	• 100 TFLOPS	• 1 PFLOPS
• DRAM, Disk, Tape (TB)	• 20, 10^3, 10^5	• 10^2, 10^4, 10^6	• 10^3, 10^5, 10^7
• Local memory Bandwidth	• 160 TBytes/s	• 800 TBytes/s	• 8 PBytes/s
Latency	• 32 ns	• 20 ns	• 15 ns
• Interconnect Bandwidth	• 20 TBytes/s	• 100 TBytes/s	• 1 PBytes/s
Latency	200 ns	125 ns	70 ns
• I/O bandwidth	• 0.6 TBytes/s	• 3 TBytes/s	• 30 TBytes/s
• Power Consumption	• 25,000 W/TFLOP	• 5,000 W/TFLOP	• 500 W/TFLOP
• Programming Model	• Hierarchical Distributed Shared Memory	• Hierarchical Distributed Shared Memory	• Distributed Shared Memory
• System Software:			
# of users supported	100s	1000s	1000s
# of threads	10^5	10^6	10^7
Sharing	static space/time	dynamic multi-user	dynamic multi-site, multi-user
Resource Management:			
Single System Image	Machine	Site	CONUS
Optimization	Manual	Tool Set	Automatic
• Programming Environment	• individual tools, run-time diagnostics, dynamic, manual parallelization tools	• fully integrated automated parallelization tools	• fully integrated automated data layout
• Security	• Strong Authentication	• Data Protection	• Multi-level Security
• Mean time between failure	• weeks	• weeks	• months